

### 11. Matching.

- a) 0.006      b) 0.777      c) - 0.923      d) - 0.487

### 13. Roller Coasters.

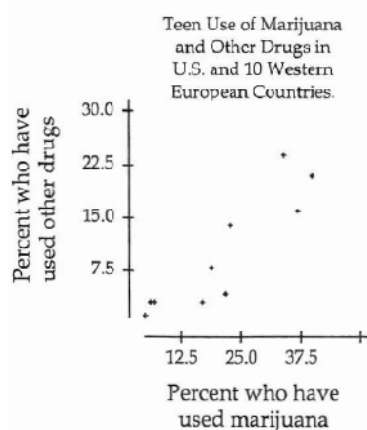
- a) It is appropriate to calculate correlation. Both height of the drop and speed are quantitative variables, the scatterplot shows an association that is straight enough, and there are not outliers.
- b) There is a strong, positive, straight association between drop and speed; the greater the height of the initial drop, the higher the top speed.

### 15. Lunchtime.

- a) The correlation between time toddlers spent at the table and the number of calories consumed by the toddlers is  $r = -0.649$ .
- b) If time spent at the table were recorded in hours instead of minutes, the correlation would not change at all. Correlation is calculated from z-scores, and since these unitless measures of position are unaffected by changes in scale, correlation is likewise unaffected.
- c) The correlation between time spent at the table and calorie consumption for toddlers is  $-0.649$ , a moderate, negative correlation. Toddlers who spent more time at the table tended to consume fewer calories.
- d) The analyst's remark is speculative. There are many possible explanations for the behavior of the toddlers. The data merely show us an association between time spent at the table and calories consumed by toddlers, and association is not the same thing as a cause-and-effect relationship.

**18. Drug abuse.**

- A scatterplot of percentage of teens who have used other drugs vs. percentage who have used marijuana in the U.S. and 10 Western European countries is at the right.
- The correlation between the percent of teens who have used marijuana and the percent of teens who have used other drugs is  $r = 0.934$ .
- The association between the percent of teens who have used marijuana and the percent of teens who have used other drugs is positive, strong, and straight. Countries with higher percentages of teens who have used marijuana tend to have higher percentages of teens that have used other drugs.
- These results do not confirm that marijuana is a “gateway drug”. An association exists between the percent of teens that have used marijuana and the percent of teens that have used other drugs. This does not mean that one caused the other.



**22. Second inning.**

- Winning teams generally enjoy greater attendance at their home games. The association between attendance and number of wins is positive, somewhat straight, but not very strong.
- The association between scoring runs and attendance has the strongest correlation, with  $r = 0.740$ .
- The correlation between number of runs scored and number of wins is  $r = 0.680$ , indicating a possible moderate association. However, since there is no scatterplot of wins vs. runs provided, we can't be sure the relationship is straight. Correlation may not be an appropriate measure of the strength of the association.

**23. Politics.**

The candidate might mean that there is an **association** between television watching and crime. The term correlation is reserved for describing linear associations between quantitative variables. We don't know what type of variables “television watching” and “crime” are, but they seem categorical. Even if the variables are quantitative (hours of TV watched per week, and number of crimes committed, for example), we aren't sure that the relationship is linear. The politician also seems to be implying a cause-and-effect relationship between television watching and crime. Association of any kind does not imply causation.

**27. Hard water.**

It is not appropriate to summarize the strength of the association between water hardness and pH with a correlation, since the association is curved, not Straight Enough.

**31. Baldness and heart disease.**

Even though the variables baldness and heart disease were assigned numerical values, they are categorical. Correlation is only an appropriate measure of the strength of linear association between quantitative variables. Their conclusion is meaningless.